DOCUMENT RESUME

ED 149 289 CS 003 861

AUTHOR Eisenberg, Anne
TITLE Lexical, Syntactic and Semantic Characteristics of the Language of Science.
PUB DATE Dec 77
NOTE 15p.; Paper presented at the Annual Meeting of the National Reading Conference (27th, New Orleans, Louisiana, December 1-3, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Content Reading; *Discourse Analysis; Readability; *Reading Comprehension; Reading Materials; *Science Instruction; *Sciences; Semantics; *Syntax

ABSTRACT

Scientific discourse is difficult to read because of its lexical, syntactic, semantic, and logical features. This paper discusses each of those features, giving special attention to syntax, and suggests ways in which discourse analysis can be used to teach content reading in sciences and to rewrite scientific text for instructional purposes. Some of the features of scientific language that can be manipulated for improved readability or that reading teachers can discuss with students to improve reading comprehension are the extensive use of passive voice, compression of abstract ideas in text without concrete examples, causal chaining, anaphora, implicit repetition, and nominalized assertions. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED149289

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Anne Eisenberg
Assistant Professor
New York City Community College
of the City University of New York
300 Jay Street
Brooklyn, New York 11201

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Anne Eisenberg

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

LEXICAL, SYNTACTIC AND SEMANTIC CHARACTERISTICS OF THE LANGUAGE OF SCIENCE

Anne Eisenberg

New York City Community College
of the
City University of New York

The discourse of science has characteristic lexical, syntactic, semantic and logical features. Determination of level of difficulty of scientific text has traditionally been approached at the lexical level; starting with Curtis (1938), reading researchers have concentrated on the vocabulary load in formulating levels of readability (Mallinson, 1972). Syntactic and semantic issues have not received the same degree of attention. They are, however, pertinent to analysis of the discourse; the teaching of reading using science text, the rewriting of such text for teaching purposes, and factors determining readability are integrally related not only to lexical features, but to syntactic, semantic and logical features of the discourse.

Lexical features include the following items: highly specialized vocabularies for each science; a lexicon characteristic of scientific English in general: consistency and rigor of terms; deliberate trimming of connotative meanings; and the use of nontechnical words with technical meanings. Syntactic features include the institutionalization of the passive; telescoping of

CS003861

information; extensive causal chaining and extensive use of anaphora, implicit repetition, and assertions which become nominalized sentences within other sentences. Semantic features include the absence of narrative voice, avoidance of figurative language and humor, and the sparing use of examples and explanations. Logical features include the use of numerical and non-numerical symbolism; evolution of a discourse which tends toward the shape provided by mathematics and formal logic; and aspects related to the discourse being essentially a written rather than verbal language.

This paper will discuss each of these features, with particular attention to syntax and recently developed computational linguistic tools available to researchers. It will also suggest ways in which characteristic sentence structures, sentence relations, and other aspects of the science register can be manipulated to develop reading skills.

Lexical Features

There are both highly specialized vocabularies for each science, and a lexicon characteristic of scientific English in general. One might expect highly specialized vocabularies within each of the sciences; since all sciences are studies of material things, a primary task is the naming of substances and objects within the area. However, it is more surprising to discover that there is a vocabulary which is common to the natural sciences, medicine, mathematics, social and behavioral sciences, political sciences, and technical and engineering literature. Yet such a lexicon exists, and has been tabulated by Zettersten in a computerized word frequency list of scientific English (1969). To a recognizable extent, there is a chemical language, a physical language, a biological language, and others within the general framework of the language of science.

A striking feature of the lexicon is the consistency of its terms. Ambiguity is the enemy, accuracy and uniformity of meaning the goal. Any particular word should have only one meaning. This meaning is sharply fixed by agreement, receiving an explicit formulation and strict adherence.

Connotative meanings are deliberately trimmed. Personal factors and all forms of affective association are avoided.

This goal is often achieved through inventing words which are not in everyday usage, and therefore not as susceptible to personal or emotive associations. For example, Faraday finished work on electrolysis and then asked Rev. William Whewell, Professor of Moral Philosophy at Cambridge, for words suitable for describing his results. That is how the terms 'anode' and 'cathode' came into existence (Savory, 1953).

Words such as 'absolute', 'critical', and 'fundamental' have precise meanings which do not vary from user to user as they do in popular language. Interestingly, they also do not vary from language to language. Science is the only arena in which translators do not have to select one word from many alternatives. There is regularly a one-for-one match between definitions in one language and another; the differences are of negligible magnitude (Bloomfield, 1939; Halliday, McIntosh and Strevens, 1964; Hertzler, 1965; Savory). For example, there is only one choice when the translator is dealing with such words as 'solution', 'precipitate', or 'crystalize'. Nybakken (1960) points out that "Many of the words used in technical English writing are identical or nearly identical in form and meaning with words used in scientific treatises written in other languages. The number of such words which are common to several languages is considerably larger in technical English than in ordinary English ...in mathematics, physics, astronomy and chemistry, the amount is large." (p. 11)

What Leonard Bloomfield called the 'translatability' of scientific discourse is one example of the consistency and rigor of the language. Another example

is found when linguists trace developments in the meaning of terms. Scientific terms do not change their meanings with the rapidity of terms in daily use. "Constancy in form and function gives to scientific words a character which distinguishes them sharply from other words, but relates them to the symbols of mathematics. Mathematic symbols such as + or $\int$ have only one possible interpretation. They have all been in existence for a long time ...Words usually change their meaning or, as Potter terms it, the penumbra of meaning. Languages grow ...The relative constancy and stability of the words of science makes [people] describe science as the enemy of language, as though it were a kind of linguistic hormone that inhibited growth and development." (Savory, p.52)

When scientists coin their own words, or have it done for them by a linguist, words can be chosen which have few or no everyday associations. However, the scientific lexicon is also made up of everyday words which have had precise meanings inflicted upon them. It is these words -- nontechnical ones which have specific meanings in their technical contexts -- which cause much difficulty for those who know only their everyday meanings. "Scientists have the outrageous habit of seizing an unsuspecting word and forcing it to do work for which it has no qualifications. The mathematic physicist is guilty of linguistic rape of a family of related words -- force, work, power and weight ...Work gives more trouble, because the physicist has decided that a force does work only when it moves something. I may push and pull in vain at some immovable obstacle, make myself hot and tired by my efforts, and yet find that mathematically I have done no work. But if I seize the dangling reins of a runaway horse and pull them, and find that nevertheless the animal continues on its course, I have had work done to me, and I,

panting and dishevelled, have done less than no work." (Savory, p. 37)

In English classes one is regularly presented with "spelling demons", words which have a history of misspelling. They have their counterpart in "meaning demons", a corpus of ostensibly non-technical words such as "absolute" , "fundamental", "uniform", and "derived". Roller (1929) points out that many words which students are expected to know from their general reading are actually technical words and need special explanation. Ordinary dictionaries are not reference books of science, and frequently contain technical definitions that are inaccurate. The following terms are examples of the kind employed without sufficient definition: absolute, critical, empirical, fundamental, intensity, modulus, rate, specific, spontaneous, theory, uniform, cohesion, conservation, dissolve, experiment, fusion, inversely, phenomenon, physical, principle, relative, symbol, unit, derive.

Lexical features of science discourse present advantages for students who are conscious of the rigor of the terminology; while one does not have the linguistic riches of ambiguity, one also does not have the ground shifting beneath one's feet. It is appropriate that lessons in vocabulary development include the difference between technical and everyday meaning of scientific terms; study skills lessons centered on the function of terms and definitions in science discourse are also pertinent.

Syntactic Features

One result of the effort to depersonalize the prose of science is the institutionalization of the passive form; it is a hallmark of the language of science.

Another characteristic of the syntax is the telescoping of information. The brevity of the language is reflected in sentence structure, and in sentence relationships. Information in science is often presented in the form of cause-effect. This is an assertion that makes sense when one considers that science is the study of the relationship of causes to effects. In expressing this relationhip in a linear way, complex causes and effects that make up a process may be pushed into a causal chain. When this happens, information is carried forward by pronominal reference, by implicit repetition, or by assertions which then become nominalized sentences within other sentences. Here is an example of one such pattern:

> The water thundering over the spillway possesses kinetic energy, which is converted to mechanical energy as it turns the blades of the turbine at the foot of the waterfall. The generator converts this mechanical energy to electrical energy, which is used to light a bulb, which in its turn gives off light energy. In this process, some of the electrical energy is also converted to heat.

If one were plotting out the information in this example as a causal chain, it might look as follows: kinetic energy of waterfall $\longrightarrow$ mechanical energy of turbine $\longrightarrow$ electrical energy of generator $\longrightarrow$ heat and light energy. In the first sentence of the example, "which" and "it" stand for "the kinetic energy of the waterfall." In the second sentence, "which", "which" and "its" stand for "electrical energy". In the

third sentence, the words "in this process" stand for the conversion of mechanical energy to electrical energy, referred to in the preceding sentence.

There are several interesting aspects of this syntactic pattern. First, the amount of anaphora or omitted words that the reader must supply is considerable. Bloomfield has commented that "In a brief speech the scientist manages to say things which in ordinary language would require a vast amount of talk ...Apparently all languages save labor by providing substitute forms whose meaning rests wholly upon the situation of speaker and hearer, especially upon earlier speech ...The most important type of substitution for our subject is anaphora; the substitute replaces repetition of a speech form which has just been uttered or is about to be uttered ...A form competent to fill one position of a construction may suffice for anaphora of a phrase embodying the whole construction." (p. 33)

Here is a second example of a causal chain. In this case, four links exist within one long sentence: The possibility that administration of digitalis, through its inhibition of the Na+K+ coupled system, produces an increase in Na+Ca++ coupled transport and thereby an increase of influx of Ca++ to the myofilaments is discussed and is presented as a possible basis for the mechanism of digitalis action.

The syntax of the sentence demands that the reader supply a considerable amount of information. This assertion can be demonstrated graphically using a device from another field, information formatting. In this field, materials are computer-processed using the natural language of scientific prose into structures called formats. Words that carry the same type of information are aligned into a single column. The columns are defined in such a way that the syntactic relations between the words of the sentence are preserved in the table. Here is a format of the sample sentence (Sager, 1977). All of the information that must be supplied by the reader is bracketed.

The possibility that administration of digitalis through its inhibition of the Na+ K+ coupled system produces an increase in Na+Ca++ coupled transport and thereby an increase of influx of Ca++ to the myofilaments is discussed and is presented as a possible basis for the mechanism of digitalis action.

HUMAN	V-STUDY	DRUG	V-CAUSE	V-QUANT	ARG1	V-PHYS	ARG2	CON
[Author]	Discusses {{ 1 2	Digitalis (Administration of)	Produces Possibly	Increase	Na+Ca++ Coupled	Transport		and thereby
		[Digitalis (Administration of)]	[Produces]	Increase	Ca++	Influx to	Myofilaments 2	through
		[Digitalis]= Its	Inhibition		Na+K+ Coupled System		1	and
[Author]	Presents	←	[{ }]				→	as basis for (possible)
		Digitalis	Action Mechanism					

This is an interesting illustration of how telescoped the language is. Note the fourth line of the format. In a chain of cause-effect, the further one goes into the chain, the more the process will be referred to pronominally or by other devices of reference. In this case, there are no physically occurring words whatsoever. The first three lines are the antecedent of as a basis for. Sager comments that, "Reasoning in science writing is characterized by devices of this sort. A single assertion becomes a nominalized sentence within another sentence, a sequence of interconnected sentences becomes an element of a later sentence by implicit repetition or by pronominal reference (this, this process, etc.)" (p. 8).

Passive verbs, embedded sentences and nominalizations are common in scientific writing. There are studies indicating that such structures present difficulties for readers. Coleman (1965) found that active verb transformations were easier to learn than their nominalizations, active versions of a sentence easier than their passive versions, and non-embedded sentences easier than their embedded counterparts. As the "simpler" sentences were often longer, Coleman concluded that readability is less effected by sentence length than by clause length. Savin (1966) found that subjects were able

to recall more words following right-branching forms than following self-embedded forms.

Awareness of syntactic aspects of science discourse has practical applications for reading teachers. To give specific instances, causal chains provide a useful note-taking technique for technical students. They are a strategy that allows students to extract information from text and set it down in an abbreviated, logical form; they are also useful for taking notes on lectures. In addition, teachers rewriting scientific text might consider avoiding passives and reducing the amount of anaphora. It is not, however, suggested that shorter sentences be substituted indiscriminately for longer ones. Often the longer sentences are the result of cause-effect relationships; chopping such sentences up into one sentence for cause and one for effect removes the explicit causal reltionship, and thereby places an inferential burden on the reader (Pearson, 1975).

The syntax of science discourse also offers certain research possibilities. Linguistic String Analysis (Sager, 1977) is a tool capable of locating and quantifying such characteristic patterns of scientific text as lexical linking and anaphora. Such analysis is pertinent to reading researchers investigating the relationship of syntactic factors to comprehension in scientific discourse.

Semantic Features

The style of scientific prose is cold, humorless and undecorated. There is a deliberate absence of narrative voice. Figurative language is shunned. Humor is absent. The language is stripped down. There is an avoidance of the use of prose to create effect. In The Double Helix, James Watson imparts the complex blending of excitement, exhaultation and envy that went into his pursuit of the structure of DNA. A biology book, it can be guaranteed, will impart nothing of

this in describing the momentous discovery. To quote Savory, "The action of a beakerful of hot nitric acid on a half pound of sugar is a fantastic sight -- but I have never seen it described in a textbook in any words more exciting than 'a copious evolution of nitrogen peroxide'."

The adage that "science is parsimonious" is reflected in the language of science. The style includes a deliberate terseness. Examples and explanations are used sparingly. From its inception, the development of scientific writing has been toward reducing the amount of words necessary for explication. The lexicon developed this way; the collapsed syntax, with its labor-saving use of pronominal and other references, tends this way; the austerity of semantic elaborations and their deliberate economy are also in this direction.

There is, in effect, a lack of redundancy, a reduction in the semantic content. This is important, because the concepts included in science are difficult. The prose is studded by first and second-order concepts. First-order concepts are induced or abstracted from concrete instances; second-order concepts are derived from two or more regular concepts (Piaget, 1963). Density, for instance, which is expressed as the ratio of mass to volume, is a second-order concept. Examples and elaborations become important in the explication of such information, as they provide a bridge for the reader from the concrete to the abstract. As Gibson and Levin point out in The Psychology of Reading, "Concepts included in the content are often necessarily abstract and difficult ...If the content of a text is difficult, leaving style out of the picture, can anything be done to assist learning from the text? One...device for improving the textual content if it is difficult it to provide examples and illustrations. It is unnecessary to cite evidence for the greater ease of comprehension of concrete than abstract material. Experiments abound, using nearly every dependent variable available, to show that abstract verbal material on almost any topic is harder to apprehend than concrete materials. The obvious implication is not to be sparing with concrete examples." (p. 421)

However, technical discourse is sparing with concrete examples. The model is one of economy in explication. The result is that the writing is saturated with concepts, with few examples or elaborations interspersed in the text. Such prose is very difficult to process. Wilson (1948) demonstrated that presenting the same number of ideas in a short passage as a longer passage, adding details, examples and explanations only, resulted in better comprehension of the ideas. Idea density has been demonstrated to be a significant influence on readability, as has degree of abstractness or concreteness (Ojemann, 1934; quoted in Kibby, 1976).

Logical Features

Both numerical and non-numerical symbolism are a part of the discourse, although both are on the edge of the language domain. The reader must switch from pose to numerical and non-numerical symbolism, and back again.

It should be noted that the language of science is essentially a written rather than verbal language, characteristically read rather than spoken. Proof of this is found in the small variation from British to American forms of scientific English, in the slowness of the change in terms over long periods of time, and in the confusion regarding pronunciation of neologisms. Savory, for instance, recounts eight possible pronunciations of the word "epigyne". Since it is rarely pronounced, no one has solved the problem (p.50).

Leonard Bloomfield gives an interesting account of how symbols which customarily are not verbally mediated may have evolved. "The ancient Greeks carried on mathematical demonstration largely in ordinary language; it was the development in the early modern period of arithmetic and algebra that divorced scientific calculations not only from ordinary language but, to all practical purposes, from vocal utterance. People learned to calculate ...by visual reception and graphic manipulation."(p.43) In the area of mathematical physics, symbols first displaced words, then made them forgettable, and finally surpassed them (Savory, p.93).

Conclusion

Lexical aspects of science discourse have been investigated (Flood, 1957; Nybukken; Zettersten) but syntax has not. Computational analysis , however, can be used to reveal characteristic syntactic patterns in scientific prose. The relationship of syntactic factors to comprehension could be looked at profitably in regard to scientific materials; the relationship of causal chains, passives and anaphora to the extraction of meaning from text are particularly pertinent. Aspects of the discourse related to its use of numerical and non-numerical symbolism, and the changes involved in going from prose to numerical symbolism, have also received little systematic investigation.

For the classroom teacher, deliberate manipulation of the conventions of science discourse can be useful. Teachers may be expected to have either little interest in science, or little confidence in their ability to handle reading lessons based on science content. In this situation, an understanding of how scientific prose works can be helpful in abating initial feelings of "science anxiety".

References

Bloomfield, L. Linguistic aspects of science. International Encyclopedia of Unified Science, 1(4), 1-59. Chicago: University of Chicago Press, 1939.

Coleman, E.B. Learning of prose written in four grammatical transformations. Journal of Applied Psychology, 1965, 49, 332-41.

Curtis, F.D. Investigations of vocabulary in textbooks of science for secondary schools. Boston: Ginn & Co., 1938.

Flood, W.E. The problem of vocabulary in the popularization of science. London: University of Birmingham Institute of Education, 1957.

Gibson, E.J. and Levin, H. The psychology of reading. Cambridge, Mass.: The MIT Press, 1975.

Hertzler, J.O. A sociology of language. New York: Random House, 1965.

Halliday, M.A.K., McIntosh, A. and Strevens, P. The linguistic sciences and language teaching. Bloomington, Indiana: Indiana University Press, 1964.

Kibby, M.W. The proper study of readability: a reaction to Carver's 'Measuring prose difficulty using the Rauding Scale'. Reading Research Quarterly, 11(4), 686-705.

Mallinson, G.G. Reading in the sciences: a review of the research. In J.L. Laffey (ed.), Reading in the content areas. Newark, Delaware: International Reading Association, 1972.

Nybakken, O.E. Greek and Latin in scientific terminology. Ames, Iowa: The Iowa State University Press, 1960.

Ojemann, R.H. The reading ability of parents and factors associated with reading difficulty of parent education materials. University of Iowa Studies in Child Welfare, 1934, 8, 11-32 (cited by Kibby, 1975)

Pearson, P.D. The effect of grammatical complexity on children's comprehension, recall and conception of certain semantic relations. Reading Research Quarterly, 10(2), 155-192.

Piaget, J. The origins of intelligence in children. International Universities Press, 1952. Reprinted by Norton, New York, 1963.

Roller, D. The terminology of physical science. Norman, Oklahoma: University of Oklahoma Press, 1929.

Sager, N. Information structures in the language of science. Paper presented at the American Association for the Advancement of Science, February 25, 1977.

Savin, H.B. Grammatical structure and the immediate recall of English sentences: 2. Embedded clauses. Cited in J. Lyonds and R.J. Wales (eds)Psycholinguistics papers. Edinburgh: Edinburgh University Press, 196

Savory, T.H. The language of science: Its growth, character and usage. London: Andre Deutsch, 1953.

Wilson, M.C. The effect of amplifying material upon comprehension. Journal of Experimental Education, Sept. 1948, 13, 5-8.

Zettersten, A. A word frequency list of scientific English. Lund, Sweden: Studentlitteratur, 1969.